Progressive Algorithm for Classifier Ensemble Construction Based on Diversity: Application to the Arabic Handwritten Recognition

Azizi Nabiha Farah Nadir Sellami Mokhtar
Labged laboratory, Badji Mokhtar University, Annaba ,Algeria
Computer Science department, Badji Mokhtar University Sidi Ammar,BO n° 12, El hadjar Annaba , 23000, Algeria
nabiha111@yahoo.fr farah@labged.net mokhtar.sellami@nasr-dz.org

***Abstract*— Multiple classifier systems (MCS) have become a popular classification paradigm for strong generalization performance. Diversity measures play an important role in constructing and explaining multiple classifier systems. The first observation concerning Arabian manuscript reveals the complexity of the task, especially for the used ensemble of classifier for recognitions. In this paper we propose a new approach Based on "Overproduce and select Paradigm" for Arabic handwritten recognition. It combines diversity measures and individual classifiers accuracy for selecting the best classifier sub. We have tested this approach with three fusion methods (Voting, Weighted voting and Bks (Behavior Knowledge Space)). The obtained experimental results are very encouraging.**

Keywords; Handwritten recognition; Ensemble classifier; Diversity measures; Overproduce and select; Classifiers fusion methods.

I. INTRODUCTION

In the field of pattern recognition, multiple classifier systems based on the combination of outputs of a set of different classifiers have been proposed as a method for the development of high performance classification systems [1]. The use of multiple classifiers as an ensemble is a promising approach to the design of reliable classifiers [2].

One of the most important tasks in optimizing a multiple classifier system is to select a group of adequate classifiers from a pool of classifiers. These methods choose a small subset from a large set of candidate models. Since there are 2^L-1 possible subsets of L models, it is not possible to try all the possibilities unless the subset L is small [4], [10]. Subset classifier selection methods also differ in the criterion they optimize. Additional to methods which directly optimize ensemble accuracy, *diversity* measures play an important role in selecting and explaining this choice of classifiers sub set [3]. In the design of ensemble classifiers, it is essential to generate a number of base classifiers with a large diversity [3,4]. Some studies rely on diversity maintenance mechanisms to generate base classifiers while others use heuristic measures to evaluate the diversity of base classifiers [5].

Off-line Handwritten Recognition remains an open problem. Studies in Arabic handwriting recognition, although not as advanced as those devoted to other scripts (e.g. Latin), have recently shown renewed interest [1, 4]. We point out that the techniques developed for Latin handwritten recognition are not appropriate for Arabic handwriting because Arabic script is based on an alphabet and rules distinct from those of Latin (cf. Section 2).

In this paper we are going to use MCS benefits for Arabic handwritten recognition. For this reason, we propose a new algorithm named *progressive algorithm* based on "Overproduce and Select paradigm" for selecting the best subset from a pool of classifiers. The proposed algorithm combines two criteria: individual classifier accuracy and diversity measures in classifier selection.

Three fusion methods are tested in this approach (voting, weighted voting, and BKS).

This paper is organized as follows: In section 2, we illustrate the related work in MCS including overproduce and select paradigm presentation, diversity measures and classifier fusion methods. In section 3, we retail the proposed approach based on Progressive algorithm. Used databases for the validation step and the experimental results are summarized in section 4.

II. Background and Related work

This section explores the current literature related to the generation of multiple classifier systems. Adopted diversity measures are also clarified after "overproduce and select" strategy presentation. We finished this section by debating the used methods of classifier fusion.

A. Related work for designing Multiple classifier systems

Roughly speaking, MCS consists of an ensemble of different classification algorithms and a decision function for combining classifier outputs. Therefore, the design of MCSs involves two main phases: the design of classifier ensemble and the design of the combination function. Although this formulation of the design problem leads one to think that effective design should address both phases, most of the design methods described in the literature focus only on the former one [6].

According to the existing literature, there are different methods to generate a MCS, all of them based on altering the training process in such way there is disagreement between the component classifiers. Different taxonomies can be considered, but it is usually agreed that there is a well known group comprising approaches considering *data resampling* to obtain different training sets to derive each individual classifier, i.e. *bagging* and *boosting*: In the bagging approach, the individual classifiers are independently learnt from resampled training sets (.bags). *Boosting* is a family of different methods following the same operation mode: the individual classifiers are generated sequentially by selecting the training set for each of them based on the performance of the previous classifier(s) in the series. These methods have gained a large acceptance in the machine learning community during the last two decades due to their high performance. Decision trees are the most usual classifier structure considered by them and much work has been done on the topic, although they can be used with any type of classifier [7].

On the other hand, a second group can be found comprised by a more diverse set of approaches which introduce the individual classifier diversity using some ways different from resampling [5]. Although some design methods have proved to be very effective and some papers have investigated the comparative advantages of different methods [2], clear guidelines are not yet available for choosing the best design method for the classification task at hand. The designer of an MCS may design a myriad of different MCSs by coupling different techniques for creating classifier ensembles with different combination functions. However, the best MCS can only be determined by performance evaluation. Accordingly, some researchers proposed the so-called "overproduce and choose" paradigm (also called "test and select" approach [7]) in order to design the MCS [7, 8, 9].

B. Overproduce and select strategy

As shown in Figure 1, the basic idea is to produce an initial large set of "candidate" classifier ensembles, and then to select the sub-ensemble of classifiers that can be combined to achieve optimal accuracy. Constraints and heuristic criteria are used in order to limit the computational complexity of the "choice" phase (e.g., the performances of a limited number of candidate ensembles are evaluated by a simple combination function like the majority voting rule [7, 8]).

- The ensemble overproduction uses techniques like Bagging and Boosting that manipulate the training set can be adopted.

Different classifiers can be also designed by using different initialisations of the respective learning parameters, using different classifier types and different classifier architectures.

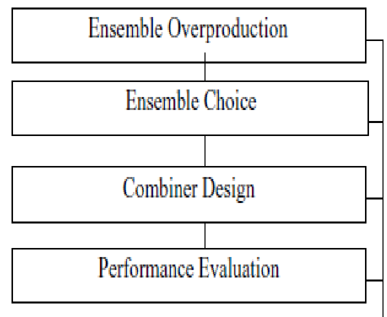

Figure 1. Overproduce and choose paradigm

The Ensemble choice phase is aimed to select the subset of classifiers that can be combined to achieve optimal accuracy.

It is easy to see that such optimal subset could be obtained by exhaustive enumeration, that is, by assessing on a validation set the classification accuracy provided by all possible subsets, and then choosing the subset exhibiting the best performance. Accordingly, techniques for evaluating the degree of error diversity of classifiers forming an ensemble have been used for classifier selection purposes.

C. Diversity In MCS selection

Diversity among the members of a team of classifiers is deemed to be a key issue in classifier combination. However, measuring diversity is not straightforward because there is no generally accepted formal

The general anticipation is that diversity measures will be helpful in designing the individual classifiers and the combination technology. Various diversity measures have been proposed in the literature [3, 4, 7]

In our approach, we have adopted three diversity measures to analyse our algorithm:

- **Correlation between the errors***:* The correlation between the classifiers errors is a natural choice to compare the classifiers subsets [3]:

$$\rho_{a,b} = \frac{Cov[v_e^a, v_e^b]}{\sqrt{Var[v_e^a]Var[v_e^b]}} \quad (1)$$

v_e^a and v_e^b are binary vectors of the *a* and *b* classifiers errors. The best set is selected while choosing the one having the minimum average of these pairs of measures.

- **Q Average:** The Q average Or Q statistic aims to calculate the similarity between two classifiers [4].

It is defined for two classifiers *a, b* as:

$$Q_{a,b} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (2)$$

Where N^{11} is the number of time where the two classifiers are correct, N^{00} the number of time where the two classifiers are incorrect.

- **Disagreement measure:** This measure represents the ratio between the number of observations where one classifier is correct and the other is incorrect [7]:

.

$$D_{a,b} = \frac{N^{10} + N^{01}}{N} \quad (3)$$

III. Proposed Approach Based on Progressive Algorithm

In overproduce phase, its uses 10 differents classifiers defined as follow:

- 02 SVM (Support Vector Machine) with the strategy "one against all ", elaborated under the library lisb SVM, version 2.7. The inputs on this SVM system are the structural features. We have used polynomial and Gaussian kernel function.

- 03 KNN (k - Nearer Neighbor: K=2 ,3,5).

- 03 NN (Neuronal Network) with different number of the hidden layer neurons and different inputs corresponding in features families detailed in section 3.

- 02 HMM (discrete and semi Continuous Hidden Marcov Models).

In the choice phase, we have observed that subset selected by only diversity measures application may not contain the most powerful classifier (with dimensions rate of recognition) or even more serious than that, can contain that the M weak classifiers which represent the most diversified ones. What inevitably degrades the recognition rate of the total system?

- For this reason, we propose the idea which tries to combine two criteria: individual ACCURACY and DIVERSITY for a subset of classifier selection.

Our proposed approach chooses a fixed M classifiers out of all L classifiers base.

1) it starts with a set containing one classifier which is the best classifier (based on individual accuracy) during the test phase ;

2) At each iteration, it chooses among all possible classifiers the one that best improves the global system performance when added to the current ensemble. The performance is calculated using evaluation criterion (the three diversity measures as we will discuss next).

Once the set of classifier is selected, it is impossible to use the methods of combination as the weighted average, or the sum of the results because outputs classifiers are heterogeneous.

Methods based on output labels classes as, voting, weight voting, and BKS will be used in our study.

IV. EXPERIMENTAL RESULTS

A. Used databases

We used two different databases in order to validate our proposed approach: the first data base representing 48 towns of Algeria names, containing 10000 words.

The second used database is The IFN/ENIT [8], database for Arabic handwritten words containing 26459 Arabic words handwritten consisting of the 946 Tunisian town/village names.

B. Experimental Approach results

In the overproduce phase: we have designed different heterogeneous classifiers indexed from 1 to10. Their individual performances using the two databases are resumed in (Table I).

In the choice phase from Overproduce and select strategy:

TABLE I. INDIVIDUAL CLASSIFIERS ACCURACY

Classifier index	**Member classifier**	**Database 01 Accuracy**	**database 02 Accuracy**
01	SVM(1)	87.18	*88.03*
02	SVM(2)	88.12	*88.39*
03	KNN(1)	74.45	*75.78*
04	KNN(2)	79.26	*79.42*
05	KNN(3)	82.62	*82.96*
06	NN(1)	84.69	*85.12*
07	NN(2)	85.08	*85.46*
08	NN(3)	85.23	*86.05*
09	HMM(1)	87.73	*88.18*
10	HMM(2)	*88.05*	*88.73*

We have executed our progressive algorithm. In this study, we have fixed M at 4 for selected set classifiers size. Three diversity measures are applied for test phase. Experimental results are resumed in table II and table III for Tunisian database and Algerian database respectively.

It is noticed that the obtained results using the progressive algorithm is better than using individual classifier separately or also the best sub set on classifiers. In any case, it is now clear that MCS performance strongly depends on careful selection of classifiers to be combined

TABLE II. BEST SUB SET CLASSIFIERS WITH THE OBTAINED PERFORMANCES

	IFN-ENIT DATABASE			
Diversity measures	**Sub set classifier**	**Vote**	**Wvot**	**BKS**
Correlation/	10, 8, 1, 9	89.18	89.56	90.11
Q Average	10, 2, 7, 8	88.15	88.78	89.25
Disagreemen t	10, 8,1, 5	89.56	90.16	90.96
Best classif	10, 2, 9,1	*88.12*	*88.68*	*89.26*

It is noticed that the obtained results using the progressive algorithm is better than using individual classifier separately or also the best sub set on classifiers. In any case, it is now clear that MCS performance strongly depends on careful selection of classifiers to be combined.

TABLE III. BEST SUB SET CLASSIFIERS WITH THE OBTAINED PERFORMANCES

	LRI- NWA DATA BASE			
Diversity measures	Sub set classifie	**Vote**	**Wvot**	**BKS**
Correlation	2, 4,9,10	89.65	89.26	90.29
Q Average	2, 1, 6, 4	88.74	88.14	89.08
Disagreeme	2, 8,10,7	89.66	90.45	90.81
Best Classif	2,1,10, 9	88.45	89.61	89.55

The effectiveness of various classifier fusion methods depends again on the selection made within classifiers.

Finally, it can be concluded that diversity is a very important factor in the subset selection of classifiers by not neglecting the individual criterion to classify accuracy.

V. CONCLUSION

In this paper, we examined the use of various diversity measures to the design of ensemble classifiers for Arabic handwritten recognition. In order to reach a decent objective, and with the optimizing aim, the progressive algorithm was proposed. Experimental results showed that the performance of designed ensemble classifiers depended on the choice of a diversity measure. Good results were obtained by the measurement of disagreement.

REFERENCE

[1] M. Cheriet, B. Horst, H. Jianying, K. Fumitaka and C.Y. Suen, “NewFrontiers in Handwriting Recognition”, Pattern Recognition,2009, Editorial

[2] L. Xu, A. Krzyzak and C.Y. Suen, “Methods of combining multiple classifiers and their applications to handwritten recognition”, IEEE trans. Syst., Man, Cybern, vol SMC-22, N°.3, pp.418-435, May/ june 1992.

[3] L. I. Kuncheva, That Elusive Diversity in Classifier Ensembles, In Proceedings of the 1st Iberian Conference on Pattern Recognition and Image Analysis, pp. 1126–1138, 2003.

[4] N. Azizi, N.Farah, M.Khadir,M.Sellami; “Arabic Handwritten Word RecognitionvUsing Classifiers Selection and features Extraction / Selection , 17 th IEEE Conf. in Inte. Info. System,IIS09, Poland, .

[5] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley–Interscience, 2004.

[6] J. Kittler, M. Hatef, R.P.W.. Duin and J. Matas, On combining classifiers, IEEE transactions on pattern analysis and mach. Intel., v. 20. 3, 1998.

[7] G. Giacinto and F. Roli (2001), An approach to the automatic design of multiple classifier systems, Patt. Recog. Le, vol 22(1), .25- 33.

[8] **Azizi N. Farah N, Sellami M, Ennaji A,** Using diversity in Classifiers set selection for Arabic handwritten recognition, 9th Work shop on Multi Classifier System, MCS’ 10, Springer, LNCS5997,pp.235-244.